Review–Hypothesis

Two functional domains conserved in major and alternate bacterial sigma factors

Patrick Stragier, Claude Parsot+ and Jean Bouvier

Institut de Microbiologie, Bât 409, Université Paris-Sud, 91405 Orsay Cedex, France and *+Unité de Biochimie Cellulaire, Institut Pasteur, 28 rue du Dr. Roux, 75724 Paris Cedex 15, France*

Received 9 May 1985

Sequences of the sigma factors of *Escherichia coli* and *Bacillus subtilis* were aligned with the sequences of two sigma-like proteins, HtpR, involved in the expression of heat-shock genes in *E. coli*, and SpoIIG, necessary for endospore formation in *B. subtilis*. An internal region is highly conserved in the four proteins and is proposed to be involved in binding of sigma factors to core RNA polymerase. The carboxy-terminal part of the four proteins presents the characteristic structure found in several prokaryotic DNA-binding proteins and is proposed to be involved in promoter recognition.

Sigma factor *RNA polymerase* *Promoter* Escherichia coli Bacillus subtilis

1. INTRODUCTION

Specificity of transcription in bacteria is dependent on the association of RNA polymerase core enzyme with a sigma factor subunit [1–3]. A single sigma factor, the *rpoD* gene product, was thought to be present in *Escherichia coli*. Recently, the product of the *E. coli htpR* gene, that controls the heat shock regulon [4], has been shown to be endowed with the properties of an authentic sigma factor [5] and to share some homology with the *E. coli rpoD* gene product [6]. During sporulation of *Bacillus subtilis* a new polypeptide is found associated with RNA polymerase that can alter its specificity in vitro [7]. This sigma-like factor could be identical or similar to the product of the *spoIIG* gene which was found to be homologous to the *E. coli* sigma factor [8]. We present a comparison of the sequences of the *E. coli* [9] and *B. subtilis* [10] major sigma factors with the sequences of the two alternate factors, HtpR and SpoIIG. This reveals striking similarities in two regions which we propose to be functional domains involved in core binding and promoter recognition.

2. A CORE-BINDING REGION

The *E. coli* HtpR protein (284 amino acids, M_r 32381) and the *B. subtilis* SpoIIG product (239 amino acids, M_r 27652) have already been compared with the *E. coli* sigma factor, σ^{70} (613 amino acids, M_r 70263) [6,8]. In both cases, a short central homologous region was found, starting at the same position, amino acid 375 in σ^{70}. The HtpR and SpoIIG sequences have now been compared with each other, and with the recently determined sequence of the *B. subtilis* major vegetative sigma factor, σ^{43} (formerly σ^{55}, 371 amino acids, M_r 42828). This comparison (fig.1) indicates that the previously observed homology is conserved in these 3 proteins and appears to extend further upstream and to involve 20 additional amino acids. A similar region can be found in σ^{70} by introducing a 245-amino acid deletion in the σ^{70} sequence from position 130 to 374 [10]. Such a large deletion could not be previously considered due to the mode of calculation chosen [11,12]. The alignment of the 4 sequences shown in fig.1 takes into account this deletion in σ^{70}.

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/85/$3.30

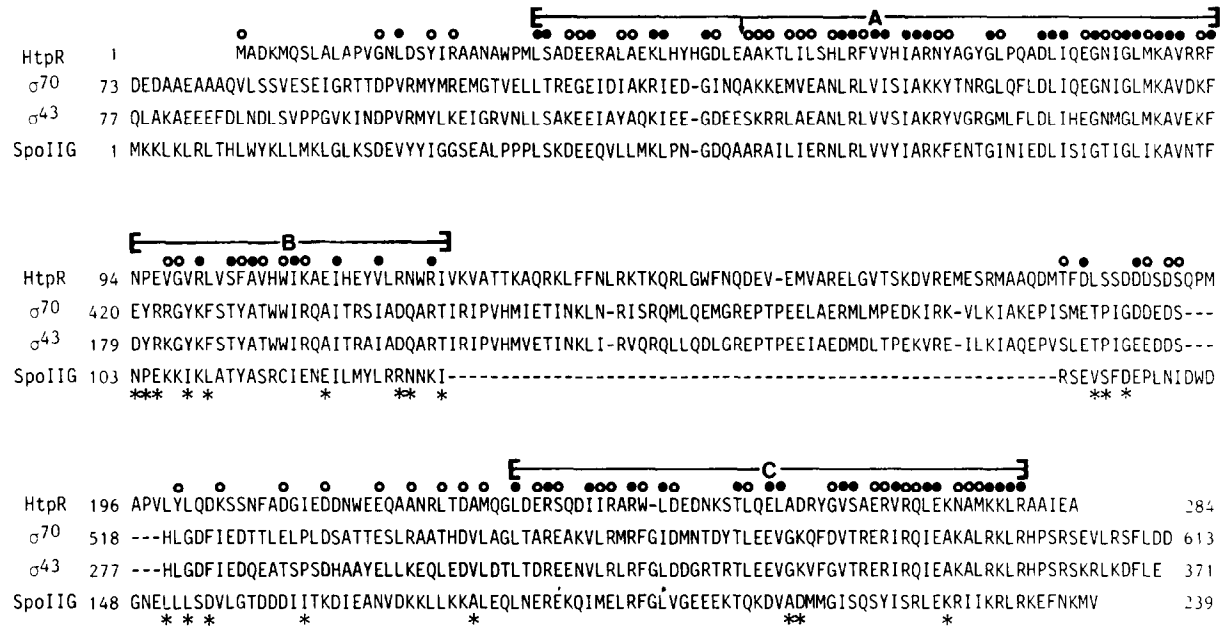

Fig.1. Alignment of the amino acid sequences of the *E. coli* HtpR and σ^{70} proteins [6,9] with the *B. subtilis* σ^{43} and SpoIIG proteins [8,10]. A 245-amino acid deletion corresponding to residues 130–374 has been introduced in the σ^{70} sequence at the position shown by the arrow. The one-letter amino acid notation is used. Chemically similar amino acids are defined as the following groups: D and E; K and R; S and T; I,L and V. Gaps have been introduced to maximize the alignments [11,12] and are shown by hyphens. Circles indicate positions occupied by identical or chemically similar amino acids in 3 (○) or 4 (●) proteins. Asterisks point to positions of homology between σ^{70} and σ^{43} on the one hand, and HtpR and SpoIIG on the other. Positions occupied by a gap in at least one of the 4 sequences were not considered. The homologous regions discussed in the text are indicated by brackets (A–C).

This representation clearly shows the existence in the 4 proteins of an internal highly conserved region (region A in fig.1) which overlaps the deletion created in σ^{70}: out of 64 positions in a row (with a single extra residue in the HtpR sequence), 45 are occupied by identical or chemically similar amino acids in at least 3 out of the 4 proteins. This high level of similarity indicates that this region must play a common and critical role for the biological activity of these 4 proteins, most probably binding to core RNA polymerase which implies very specific amino acid interactions. *B. subtilis* and *E. coli* core RNA polymerases are very similar in size and subunit composition [13]; moreover, they can promote specific transcription in vitro with each other's sigma factor [14,15], which demonstrates a degree of homology sufficient to allow heterologous interactions. Furthermore, careful examination of the adjacent downstream sequences reveals a quite unexpected pattern (see fig.1, region B): two homology subclasses can be observed, not matching the bacterial origin of the proteins but gathering the major sigma factors on the one hand and the alternate proteins, HtpR and SpoIIG, on the other. In 9 out of 30 positions, pairs of identical or chemically similar amino acids can be related to each subclass while 7 positions are common to both. This structural relationship may be the reflection of a similar functional role: both HtpR and the sporulation sigma-like factors (SpoIIG among others) must displace the preexisting major sigma factor in order to allow proper genetic adaptation to the external changes in environment (heat shock or nutrient deprivation). In vitro experiments indicate that these displacements do not require any prior chemical modification of the core RNA polymerase, but seem to result from mere competition between both kinds of proteins [5,7]. This implies that the alternate sigma factors have

a higher affinity for the core enzyme than the major sigma factors, which is in good accordance with their different behaviour during chromatography purification steps [5,7]. A stronger binding to the core enzyme could be achieved by additional interactions due to the presence of proper amino acids in the immediate vicinity of the major binding domain. The amino acids pointed out above as specific to HtpR and SpoIIG could play this role.

The major homology region found in HtpR and in the two *B. subtilis* factors, σ^{43} and SpoIIG, is not interrupted while it is split into two domains by a 245-amino acid block in σ^{70}. This suggests that duplication of an ancestral sigma gene preceded divergence between *E. coli* (a Gram-negative bacterium) and *B. subtilis* (Gram-positive) and that the present structure of σ^{70} could be the result of a fusion event by insertion of an extra coding sequence.

3. A PROMOTER-RECOGNITION REGION

Several equally possible alignments can be found for the next 100 positions, all accommodating a 58-amino acid gap in the SpoIIG sequence and a few additional residues in the HtpR protein. One of these alignments has been represented in fig.1. Some homologies can still be observed but, as they depend on the choice of the alignment, their significance appears too weak to be further discussed. Conversely, a 50-amino acid domain is highly conserved in the carboxy-terminal part of the 4 proteins (see fig.1, region C). As already noticed for the HtpR protein [6], this region shares all the properties of the DNA-binding domains of several prokaryotic repressors and activators [16]. A similar α-helix–β-turn–α-helix structure can be predicted since the amino acids important for such a conformation are conserved as detailed in fig.2a. There is only one exception: in the σ^{70} sequence the conserved glycine in the β-turn is replaced by aspartic acid. However, a glutamic acid residue can be accommodated at this position in a functional mutated λ repressor [17]. Such a structure strongly suggests that this region of the sigma factors might bind to DNA and be involved in promoter recognition. Although σ^{70} is not a DNA-binding protein per se [18], crosslinking experiments have demonstrated that, when complexed with the core enzyme, it is in close contact with nucleotides located in the promoter region [19,20]. As exemplified by mutations affecting recognition properties of the cAMP receptor protein in *E. coli* [21], the specificity of DNA binding appears to be determined by the side groups protruding from the

		HELIX								HELIX									
		P	P	P	H	Gly/Ala	P	-	H	Gly	H	P	P	P	P	Ile/Val	P	P	H	-	P
a																					
HtpR	253	Leu	Gln	Glu	Leu	Ala	Asp	Arg	Tyr	Gly	Val	Ser	Ala	Glu	Arg	Val	Arg	Gln	Leu	Glu	Lys 272
σ^{70}	573	Leu	Glu	Glu	Val	Gly	Lys	Gln	Phe	Asp	Val	Thr	Arg	Glu	Arg	Ile	Arg	Gln	Ile	Glu	Ala 592
σ^{43}	332	Leu	Glu	Glu	Val	Gly	Lys	Val	Phe	Gly	Val	Thr	Arg	Glu	Arg	Ile	Arg	Gln	Ile	Glu	Ala 351
SpoIIG	206	Gln	Lys	Asp	Val	Ala	Asp	Met	Met	Gly	Ile	Ser	Gln	Ser	Tyr	Ile	Ser	Arg	Leu	Glu	Lys 225
b																					
HtpR	154	Val	Glu	Met	Val	Ala	Arg	Glu	Leu	Gly	Val	Thr	Ser	Lys	Asp	Val	Arg	Glu	Met	Glu	Ser 173
σ^{70}	480	Pro	Glu	Glu	Leu	Ala	Glu	Arg	Met	Leu	Met	Pro	Glu	Asp	Lys	Ile	Arg	Lys	Val	Leu	Lys 499
σ^{43}	239	Pro	Glu	Glu	Ile	Ala	Glu	Asp	Met	Asp	Leu	Thr	Pro	Glu	Lys	Val	Arg	Glu	Ile	Leu	Lys 258
SpoIIG	154	Ser	Asp	Val	Leu	Gly	Thr	Asp	Asp	Asp	Ile	Ile	Thr	Lys	Asp	Ile	Glu	Ala	Asn	Val	Asp 173

Fig.2. Alignment of the 4 sigma factors showing the regions proposed to interact with DNA. The highly conserved carboxy-terminal domains are indicated in part a, while the weakly homologous internal domains are shown in part b. The residues conserved in other prokaryotic DNA-binding proteins [16] are indicated (H, hydrophobic; P, polar). The position of the α-helix–β-turn–α-helix corresponding structure is shown.

second α-helix which lies in the major groove. This region contains the largest uninterrupted stretch of identical amino acids found in σ^{70} and σ^{43}, which are known to recognize the same consensus promoter sequences [22]. On the other hand, HtpR allows recognition of specific heat shock promoters of distinct sequences [23] and sporulation-specific promoters are quite different from vegetative ones in *B. subtilis* [24]; several distinct amino acids are present in this region of the HtpR and SpoIIG proteins although they can lead potentially to similar secondary structures.

If the carboxy-terminal part of the sigma factors is involved in promoter recognition, mutations altering transcriptional specificity are expected to be found in this region. Such mutations have already been isolated in *E. coli* and provided the first genetic evidence for the role of σ^{70} in promoter selection [25]. They have been located in Arg-596 that lies 4 residues downstream of the second α-helix described above [26]. Their properties can be interpreted either as affecting the conformation of the adjacent 'recognition helix' [16] or as modifying specific contacts with the ancillary activator proteins, AraC and MalT. The analysis of other similar mutations should give more information on the recognition role of this region.

As prokaryotic promoters appear to be constituted of two separate regions, the existence of two distinct DNA-recognition sites in the sigma factors was postulated [3]. A second putative DNA-binding region was proposed for the HtpR protein at positions 154–173 [6]. A similar structure may also exist in the related regions of σ^{70} and σ^{43} as detailed in fig.2b, but only a weak equivalent can be found in SpoIIG at a different location. Although less conserved than the carboxy-terminal domains, this region could also be involved in promoter selectivity. A schematic representation of the 4 sigma proteins and of their conserved regions is given in fig.3.

4. A GENERAL MODEL?

The homologies found in two regions of sigma factors belonging to evolutionarily distant bacteria are strong indications of their functional importance. The sequences proposed to be part of the core binding region are not present in the *E. coli nusA* gene product, a protein involved in termination and antitermination of transcription and able to bind to core polymerase in vitro [27]. This suggests that initiation and termination factors interact with different regions of core RNA polymerase. The sigma factors encoded by gene *28* and genes *33* and *34* from *B. subtilis* bacteriophage SP01 and by gene *55* from *E. coli* bacteriophage T4 do not contain the proposed core binding sequence [28–30] although they are well documented to be able to reprogram unmodified host core polymerase for recognition of bacteriophage specific promoters [31,32]. Due to their viral origin these proteins may have acquired their properties through convergent evolution and use different amino acids to achieve similar tertiary structures.

The existence of two distinct highly conserved

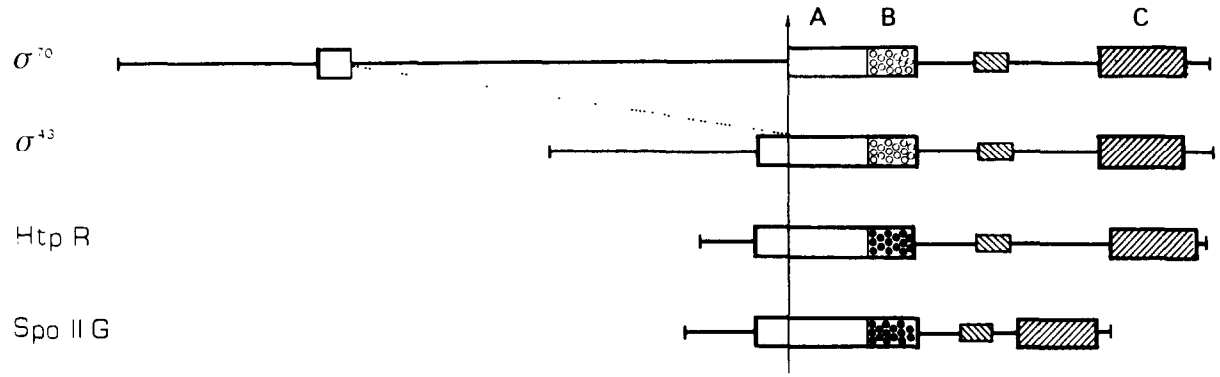

Fig.3. Schematic representation of the 4 sigma proteins. The two highly conserved domains (regions A–B and C) are shown by large heavy-lined boxes and the possible second DNA-binding sites by small light-lined boxes. The core-binding region is shown as an open box (split into two parts in σ^{70}) adjacent to an ancillary domain that may enhance core binding in HtpR and SpoIIG (●) as compared to σ^{70} and σ^{43} (○). The DNA-binding regions are hatched.

domains in bacterial sigma factors is in good accordance with their bifunctional role. Interaction with core polymerase through a common region might be modulated by the presence of adjacent variable sites while promoter recognition would involve a widely used DNA-binding structure [16]. The organization described in fig.3 is expected to be found in other alternate sigma factors, for instance in the minor vegetative *B. subtilis* sigma factors, σ^{28}, σ^{32}, σ^{37} [3] and their *Streptomyces coelicolor* counterparts [33], and also in as yet uncharacterized proteins required for the expression of complex gene sets. For instance, *Klebsiella pneumoniae* and *Rhizobium* genes involved in nitrogen assimilation share a quite specific promoter sequence [34]; they are controlled by the *nifA* and *ntrA* gene products which were proposed to be specific sigma factors [35,36]. Sequencing of these genes could give a first answer.

ACKNOWLEDGEMENTS

We are grateful to R.H. Doi for friendly collaboration and generous sharing of unpublished data. We thank D. Dean for critical review of the manuscript.

REFERENCES

[1] Travers, A.A. and Burgess, R.R. (1969) Nature 222, 537–540.
[2] Chamberlin, M.J. (1976) in: RNA Polymerase (Losick, R. and Chamberlin, M.J. eds) pp.159–191, Cold Spring Laboratory, Cold Spring Harbor, NY.
[3] Losick, R. and Pero, J. (1981) Cell 25, 582–584.
[4] Neidhart, F.C., VanBogelen, R.A. and Lau, E.T. (1983) J. Bacteriol. 153, 597–603.
[5] Grossman, A.D., Erickson, J.W. and Gross, C.A. (1984) Cell 38, 383–390.
[6] Landick, R., Vaughn, V., Lau, E.T., VanBogelen, R.A., Erickson, J.W. and Neidhart, F.C. (1984) Cell 38, 175–182.
[7] Haldenwang, W.G., Lang, N. and Losick, R. (1981) Cell 23, 615–624.
[8] Stragier, P., Bouvier, J., Bonamy, C. and Szulmajster, J. (1984) Nature 312, 376–378.
[9] Burton, Z., Burgess, R.R., Lin, J., Moore, D., Holder, S. and Gross, C.A. (1981) Nucleic Acids Res. 9, 2889–2903.
[10] Gitt, M.A., Wang, L.F. and Doi, R.H. (1985) J. Biol. Chem., in press.
[11] Needleman, S.B. and Wunch, C.D. (1970) J. Mol. Biol. 48, 443–453.
[12] Smith, T.F., Waterman, M.S. and Fitch, W.M. (1981) J. Mol. Evol. 18, 38–46.
[13] Losick, R., Shorenstein, R.G. and Sonenshein, A.L. (1970) Nature 227, 910–913.
[14] Shorenstein, R.G. and Losick, R. (1973) J. Biol. Chem. 248, 6170–6173.
[15] Achbeyer, E.C. and Whiteley, H.R. (1980) J. Biol. Chem. 255, 11957–11964.
[16] Pabo, C.O. and Sauer, R.T. (1984) Annu. Rev. Biochem. 53, 293–321.
[17] Hochschild, A., Irwin, N. and Ptashne, M. (1983) Cell 32, 319–325.
[18] Zillig, W., Zechel, K., Rabussay, D., Schachner, M., Sethi, V., Palm, P., Heil, A. and Seifert, W. (1970) Cold Spring Harb. Symp. Quant. Biol. 35, 47–58.
[19] Simpson, R.B. (1979) Cell 18, 277–285.
[20] Chenchick, A., Beabealashvilli, R. and Mirzabekov, A. (1981) FEBS Lett. 128, 46–50.
[21] Ebright, R.H., Cossart, P., Gicquel-Sanzey, B. and Beckwith, J. (1984) Nature 311, 232–235.
[22] Moran, C.P. jr, Lang, N., LeGrice, S.F.J., Lee, G., Stephens, M., Sonenshein, A.L., Pero, J. and Losick, R. (1982) Mol. Gen. Genet. 186, 339–346.
[23] Neidhardt, F.C., VanBogelen, R.A. and Vaughn, V. (1984) Annu. Rev. Genet. 18, 295–329.
[24] Johnson, W.C., Moran, C.P. jr and Losick, R. (1983) Nature 302, 800–804.
[25] Travers, A.A., Buckland, R., Goman, M., LeGrice, S.S.G. and Scaife, J.G. (1978) Nature 273, 354–358.
[26] Hu, J.C. and Gross, C.A. (1985) Mol. Gen. Genet. 199, 7–13.
[27] Ishii, S., Ihara, M., Maekawa, T., Nakamura, Y., Uchida, H. and Imamoto, F. (1984) Nucleic Acids Res. 12, 3333–3342.
[28] Costanzo, M. and Pero, J. (1983) Proc. Natl. Acad. Sci. USA 80, 1236–1240.
[29] Costanzo, M., Brzustowicz, L., Hannett, N. and Pero, J. (1984) J. Mol. Biol. 180, 533–547.
[30] Gram, H. and Rüger, W. (1985) EMBO J. 4, 257–264.
[31] Talkington, C. and Pero, J. (1978) Proc. Natl. Acad. Sci. USA 75, 1185–1189.
[32] Kassavetis, G.A. and Geiduschek, E.P. (1984) Proc. Natl. Acad. Sci. USA 81, 5101–5105.
[33] Westpheling, J., Ranes, M. and Losick, R. (1985) Nature 313, 22–27.
[34] Ausubel, F.M. (1984) Cell 37, 5–6.
[35] De Bruijn, F.J. and Ausubel, F.M. (1983) Mol. Gen. Genet. 192, 342–353.
[36] Raibaud, O. and Schwartz, M. (1984) Annu. Rev. Genet. 18, 173–206.